

RESEARCH

Open Access



Genetic risk factors modulate the association between physical activity and colorectal cancer

Anita R. Peoples^{1,2,3†}, Mireia Obón-Santacana^{4,5,6†}, Andre E. Kim⁷, Eric S. Kawaguchi⁷, Yubo Fu⁷, Conghui Qu⁸, Ferran Moratalla-Navarro^{4,5,6,9}, John Morrison⁷, Yi Lin⁸, Volker Arndt¹⁰, Sonja I. Berndt¹¹, Stephanie A. Bien⁸, D. Timothy Bishop¹², Emmanouil Bouras¹³, Hermann Brenner^{10,14,15}, Daniel D. Buchanan^{16,17,18}, Peter T. Campbell¹⁹, Andrew T. Chan^{20,21,22,23,24,25}, Jenny Chang-Claude^{26,27}, David V. Conti⁷, Douglas A.C. Corley^{28,29}, Matthew A. Devall^{30,31}, Niki Dimou³², David A. Drew^{20,22}, Stephen B. Gruber^{33,34}, Marc J. Gunter^{32,35}, Sophia Harlid³⁶, Tabitha A. Harrison⁸, Michael Hoffmeister¹⁰, Li Hsu^{8,37}, Jeroen R. Huyghe⁸, Temitope O. Keku³⁸, Anshul Kundaje^{39,40}, Juan Pablo Lewinger⁷, Li Li^{41,42}, Brigid M. Lynch^{43,44}, Loic Le Marchand⁴⁵, Vicente Martín^{6,46,47}, Neil Murphy³², Christina C. Newton³, Shuji Ogino^{23,48,49}, Sheetal Hardikar^{1,2}, Jennifer Ose^{1,2,50}, Rish K. Pai⁵¹, Julie R. Palmer⁵², Nikos Papadimitriou³², Bens Pardamean⁵³, Andrew J. Pellatt⁵⁴, Mila Pinchev⁵⁵, Elizabeth A. Platz⁵⁶, John D. Potter^{8,57,58}, Gad Rennert^{59,60}, Edward A. Ruiz-Narvaez⁶¹, Lori C. Sakoda^{8,28}, Robert E. Schoen⁶², Anna Shcherbina^{39,40}, Mariana C. Stern⁶³, Yu-Ru Su⁷⁰, Claire E. Thomas⁸, Yu Tian^{26,64}, Konstantinos K. Tsilidis^{13,35}, Caroline Y. Um³, Franzel J. B. van Duijnhoven⁶⁵, Bethany Van Guelpen^{36,66}, Kala Visvanathan⁵⁶, Jun Wang^{63,67}, Emily White^{8,55}, Alicja Wolk⁶⁸, Michael O. Woods⁶⁹, Anna H. Wu⁶³, Cornelia M. Ulrich^{1,2*†}, Ulrike Peters^{8,58*†}, W. James Gauderman^{7*†} and Victor Moreno^{4,5,6,9*†}

Abstract

Background Physical activity is an established protective factor for colorectal cancer (CRC), but it is unclear if genetic variants modify this effect. To investigate this possibility, we conducted a genome-wide gene–physical activity interaction analysis.

Methods Using logistic regression (1-d.f), two-step screening and testing method (EDGE), and joint tests (3-d.f), we analyzed interactions between common genetic variants across the genome and physical activity in relation to CRC

†Anita R. Peoples and Mireia Obón-Santacana contributed equally and are joint 1st authors.

†Cornelia M. Ulrich, Ulrike Peters, W. James Gauderman, and Victor Moreno contributed equally and are joint last authors and corresponding authors.

*Correspondence:

Cornelia M. Ulrich
neli.ulrich@hci.utah.edu

Ulrike Peters
upeters@fredhutch.org

W. James Gauderman
jim@usc.edu

Victor Moreno
v.moreno@iconcologia.net

Full list of author information is available at the end of the article



risk. Self-reported physical activity levels were categorized as active (≥ 8.75 MET-h/wk) vs. inactive (< 8.75 MET-h/wk; 39,992 participants) and as study- and sex-specific quartiles of activity (42,602 participants).

Results Physical activity was inversely associated with CRC risk overall (OR [active vs. inactive] = 0.85; 95% CI = 0.81–0.90). The two-step EDGE method identified an interaction between rs4779584, an intergenic variant near the *GREM1* and *SCG5* genes, and physical activity for CRC risk (p -interaction = 2.6×10^{-8}). Stratification by genotype at this locus showed a significant reduction in CRC risk by 20% in active vs. inactive participants with the CC genotype (OR = 0.80; 95% CI = 0.75–0.85), but no significant physical activity–CRC associations among CT or TT carriers. When physical activity was modeled as quartiles, the 1-d.f. test identified that rs56906466, an intergenic variant near the *KCNQ1* gene, modified the association between physical activity and CRC (p -interaction = 3.5×10^{-8}). Stratification at this locus showed that an increase in physical activity (highest vs. lowest quartile) was associated with a lower CRC risk solely among TT carriers (OR = 0.77; 95% CI = 0.72–0.82).

Conclusions In summary, we identified two genetic variants that modified the association between physical activity and CRC risk. One of them, related to *GREM1* and *SCG5*, suggests that the bone morphogenetic protein (BMP)-related, inflammatory, and/or insulin signaling pathways may be involved in the protective association between physical activity and colorectal carcinogenesis.

Keywords Physical activity, Gene-environment interaction, Colorectal cancer, GWAS

Background

Colorectal cancer (CRC) is a major global cause of morbidity and mortality. It is the third most commonly diagnosed cancer and the second leading cause of death in the world, with more than 1.9 million incident cases and 0.9 million deaths in 2022 [1]. It is predicted that there will be 2.2 million and 3.2 million new CRC cases by 2030 [2] and 2040 [3], respectively, confirming CRC as a major continuing public health burden. The underlying etiology of CRC is multifactorial, with a combination of genetic and environmental factors increasing the likelihood of developing CRC. Among these risk factors, physical activity, a lifestyle factor, is an established protective factor against CRC [4–7].

Multiple observational studies and several systematic reviews have shown that regular physical activity (occupational or leisure time) is a modifiable factor associated with lower CRC risk [8–11]. In particular, the World Cancer Research Fund/American Institute for Cancer Research (WCRF/AICR) Continuous Update Project reported lower CRC risk with increased physical activity and classified the evidence linking physical activity to lower CRC risk as “strong” [4]. Despite the beneficial health effects of physical activity, more than a quarter of all adults globally do not engage in sufficient physical activity [12].

There is substantial understanding of the mechanisms underlying the protective association of physical activity with CRC risk; for example, physical activity is known to have beneficial effects on skeletal muscle mass, immune function, sleep, and mental health [5, 13–17]. Physical activity also reduces obesity (fat mass), which has a beneficial effect on CRC through a

reduction in insulin resistance and inflammation, both of which have been associated with CRC development [5, 18, 19]. More recently, physical activity has been linked to improved gut microbiome diversity [20, 21]. Further, genetic factors may play a role in modifying the relationship between physical activity and CRC, as investigated in some gene–environment (GxE) interaction studies [22–26]. However, most previous studies have focused on candidate gene approaches or pathway-based frameworks, or were limited by small sample sizes and thus underpowered to detect genome-wide significant interactions [22–25]. A recent genome-wide interaction study by Cho et al. [26] using UK Biobank data evaluated physical activity–gene interactions in relation to CRC risk among 2979 CRC cases and 11,435 matched controls. However, even in this relatively large-scale, genome-wide analysis, no SNP, gene, or pathway reached statistical significance after correction for multiple testing, underscoring the ongoing challenges in detecting robust GxE interactions for complex traits.

Understanding the genetic factors that may influence the relationship between physical activity and CRC risk can offer novel insights into potential biological mechanisms of colorectal carcinogenesis, as well as better inform efforts to promote physical activity and potentially identify individualized physical activity prescriptions. We conducted the largest genome-wide GxE analysis to date, aiming to identify novel genetic variants that may modify the protective association between self-reported physical activity and CRC risk in order to obtain insight into potential mechanisms behind this association.

Methods

Study participants

The study included individual-level genomic and epidemiologic data from three CRC consortia: the multi-centered Colon Cancer Family Registry (CCFR), the Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO), and the Colorectal Cancer Transdisciplinary Study (CORECT), which have been previously described [27–31]. Nested case–control sets were assembled from cohort studies. Control participants were matched on age, sex, and enrollment date/trial group, when applicable. CRC cases were defined as invasive colon or rectal tumors and were confirmed via multiple sources including electronic medical records, pathology reports, state or provincial cancer registries, and/or death certificates. For the small subset of advanced adenomas (7–8%), matched controls were polyp-free and were confirmed by sigmoidoscopy or colonoscopy at the time of adenoma diagnosis. Each study was approved by relevant ethics committees or review boards from respective institutions. All participants provided written informed consent at recruitment.

Data harmonization

All data, including physical activity, were collected, centralized, and harmonized at the GECCO consortium coordinating center at the Fred Hutchinson Cancer Center using a standardized protocol to ensure that all variables were comparable across studies [31]. Briefly, data harmonization consisted of a multi-step procedure that reconciled differences in study protocols and data-collection instruments [32]. Common data elements (CDEs) were defined a priori for harmonization. Study questionnaires and data dictionaries for each study were reviewed to identify study-specific data elements, which were then mapped to the CDEs through an iterative process of communication with data contributors to obtain relevant data and coding information. These data elements were transformed and integrated into a single unified database with standardized definitions, coding, and permissible values, implemented via SAS and T-SQL. Resulting data were checked for quality assurance, errors, and outlying values within and between studies [33]. Outliers were truncated to the minimum or maximum value of the established range for each variable.

Epidemiologic and lifestyle data collection

Information on demographic, lifestyle, and environmental factors as well as potential risk factors such as age at diagnosis or enrollment, sex, education level, smoking status, total energy consumption (kcal/day), and self-reported or measured weight and height

were collected via in-person interviews or through structured self-administered questionnaires in each study. Total energy consumption was derived from the Food Frequency Questionnaires, with missing values imputed by study-sex-specific means. Body mass index (BMI) was calculated using the weight (kg) and height (m) of each participant.

Physical activity exposure measure

Information on physical activity was obtained from structured, study-specific questionnaires, such as the International Physical Activity Questionnaire (IPAQ) short form [34], European Prospective Investigation into Cancer and Nutrition (EPIC) physical activity questionnaire, and Nurses' Health Study physical activity questionnaire, among others. These instruments were either self-administered or completed via in-person interviews. Physical activity was assessed at the study reference time, which was defined as study entry (start of follow-up) in cohort studies and the period 1 to 2 years prior to study entry for case–control studies, to ensure physical activity was measured prior to cancer diagnosis. Detailed information on physical activity data collection is provided in Additional file 1: Table 1. To harmonize physical activity across studies, we defined total physical activity as the sum of leisure-time and, when assessed, other reported activity domains (e.g., occupational or transportation-related), expressed in metabolic equivalent task hours per week (MET-h/wk). This harmonized measure was derived for each participant to reflect the approximate average time per week that the individual spent in leisure activities and/or all reported activities.

Moderate activity was defined as 3.5 to 6 MET-h/wk and vigorous activities as ≥ 6 MET-h/wk [35]. Thus, at least 8.75 MET-h/wk approximately corresponds to the current physical activity guidelines of a minimum of 150 min (= 2.5 h) of moderate or 75 min of vigorous activity per week as recommended for individuals with cancer or for cancer prevention [36–39]. Based on these guidelines and previously published literature in CRC [40–42], the participants in the present study were categorized into two groups: active (≥ 8.75 MET-h/wk) vs. inactive (< 8.75 MET-h/wk; reference category). Because the majority of the participants were active, we also calculated study- and sex-specific quartiles for physical activity as a secondary variable, where the quartile groups were coded as 1, 2, 3, or 4, respectively. This variable was treated as continuous (change in one quartile) when assessing the association between physical activity and CRC, and as categorical (1st quartile as reference group) in the genome-wide scans.

Genotyping, quality control, and imputation

Detailed information on genotyping, imputation, and quality control has been described previously [27, 29]. In brief, genotyped single nucleotide polymorphisms (SNPs) were excluded based on deviation from Hardy–Weinberg Equilibrium ($p < 1 \times 10^{-4}$), low call rate ($< 95\text{--}98\%$), discrepancies between reported and genotypic sex, and discordant calls between duplicates. Autosomal SNPs in all studies were imputed to the Haplotype Reference Consortium (HRC) r1.1 (2016) panel using the University of Michigan Imputation Server [43] and treated as dosage for data management and analyses using R package BinaryDosage [44]. Imputed common SNPs were excluded if they had low imputation quality ($R^2 < 0.8$) and pooled minor allele frequency (MAF) $\leq 1\%$. After quality control, a total of over > 7.2 million SNPs were used for the gene–environment interaction analysis, noticeably with high redundancy due to linkage disequilibrium (LD).

Sample size

Analyses were limited to individuals of European ancestry, based on self-reported race and clustering of principal components (PCs) with 1000 Genomes EUR superpopulations [45]. Participants were excluded based on cryptic relatedness or duplicates (prioritizing cases and/or individuals genotyped on the better platform), genotyping/imputation errors, and extreme outlier values for physical activity. We also excluded studies that did not collect physical activity data, had high proportions of missing values for physical activity, used case-only designs, or exhibited implausibly wide distributions of physical activity that could not be harmonized with other studies. After these exclusions, the final pooled sample size for the study- and sex-specific quartile physical activity variable was 42,602 participants from 31 studies (71% prospective cohort studies). For the dichotomous active vs. inactive physical activity variable, with 8.75 MET-h/wk as the cutoff value, the final pooled sample size was 39,992 participants from 27 studies (74% prospective cohort studies; Additional file 1: Table 1).

Statistical analyses

To evaluate the main effects of physical activity on CRC risk, logistic regression models were conducted for each study, with adjustment for age at diagnosis or enrollment, sex, and total energy consumption (when available). Models with genetic variables were further adjusted for the first three PCs of genetic ancestry to account for potential population substructure. The study-specific results were combined using random-effects meta-analysis methods (Hartung–Knapp) to obtain summary odds ratios (ORs) and 95% confidence intervals (CIs) [46]. The heterogeneity p -values were calculated using Cochran's Q

statistics [47], while funnel plots identified studies with outlying ORs for potential exclusion and sensitivity analyses. Additional models were fitted, stratified by study design (case–control vs. cohort), sex, and tumor site (proximal colon, distal colon, rectal). All meta-analyses were performed using the R package Meta [48].

Genome-wide interaction scans of common markers were conducted in the overall study population to maximize power. For the purposes of this study, E indicates physical activity, G indicates a particular SNP, D indicates CRC disease status, and C refers to a set of adjustment covariables. We utilized not only the traditional logistic regression test of Gx E (1-degree of freedom test; 1-d.f.), but also the more powerful joint 3-d.f. test [49, 50] and two-step screening and testing method (EDGE) method [51–53]. The R package GxEScanR [54] was used to perform these analyses.

For the 1-d.f. test, we examined multiplicative interactions by fitting a traditional logistic regression model including an interaction term in the form: $\text{logit}(\text{Pr}(D = 1|G)) = \beta_0 + \beta_G G + \beta_E E + \beta_{GxE} GxE + \beta_C C$, where $H_0: \beta_{GxE} = 0$ tests potential departures from multiplicative associations of E and G on D .

We also performed a joint test of association, which can improve power to detect disease susceptibility loci in a wider range of circumstances by accounting for Gx E interactions, e.g., in circumstances where susceptibility loci affect only individuals with certain environmental exposure profiles [50, 55]. For this we used the 3-d.f. test of the joint null hypothesis $H_0 = \beta_G = \beta_{GxE} = \gamma_G = 0$, where β_G and β_{GxE} are the main and interaction effects from the logistic model above and γ_G represents the association between G and E in the combined case–control sample [50, 56].

We further implemented the two-step EDGE method that assesses Gx E interaction tests (step 2) based on ranks of an independent filtering or ranking statistic (step 1) [53]. The two-step method can decrease the multiple testing burden and improve power to detect interaction loci [53, 56, 57], provided that steps 1 and 2 are independent. The original approach uses step 1 ranks to prioritize and partition SNPs into exponentially larger bins of fixed sizes and increasingly more stringent step-2 significance thresholds. However, when analyzing imputed SNPs, highly correlated markers from the same loci fill the top bins, thereby diminishing statistical power. To address this issue, the original weighted hypothesis-testing framework [58] was modified to accommodate bins of varying sizes while appropriately controlling for type I error [52]. In particular, SNPs were partitioned into bins based on step 1 p -value thresholds *in expectation*, which were calculated using the original predetermined bin sizes (initial bin size of 5 and overall $\alpha = 0.05$) with

assumed uniform distribution of 1 million independent tests. For step 2 GxE testing, the influx of correlated markers into each bin was accounted for by correcting for the effective number of tests, which was estimated using principal component analysis (PCA) performed on bin-specific genotype correlation matrices [51, 52, 59]. This modification reduces multiple testing burden and improves statistical power, while preserving the overall type I error rate at 5%. For any SNP achieving significance at the overall type I error rate, we computed its corresponding SNP-specific p-value accounting for both steps 1 and 2 of the EDGE procedure, to allow direct comparison to the standard GWAS threshold of 5×10^{-8} [59].

To follow up statistically significant interactions, we estimated stratified ORs by modeling physical activity in relation to CRC within genotypic groups and the per-allele increase in genotype in relation to CRC stratified by physical activity. We also assessed the extent of genomic inflation by creating quantile–quantile (Q-Q) plots and calculating the genomic inflation factor (λ). Additionally, we calculated λ_{1000} , which scales the genomic inflation factor to an equivalent study of 1000 cases and 1000 controls, since λ scales according to the sample size [60, 61].

To explore variation in GxE effect strengths of association, we also conducted stratified analyses for novel findings by study design, sex, and tumor site. We conducted a sensitivity analysis including the interaction terms GxBMI and E(=physical activity)xBMI in the model, because BMI is a potential confounder in the physical activity–CRC association [62].

Functional follow-up

Regional plots for all statistically significant findings were generated using the command-line version (standalone) of LocusZoom v1.3 [63] to examine, in depth, the magnitudes of association, the extent of association signal due to LD, and chromosomal position of findings relative to genes in the given region. Measures of LD were estimated using study population controls. The putative functional role of these SNPs and those in LD ($R^2 > 0.5$) at 500 kb flanking regions were examined relative to their potential contribution to regulate gene expression by their (i) direct association with expression of nearby genes (expression quantitative trait loci (eQTLs)) and (ii) physical location in regions of chromatin accessibility or histone modifications (variant enhancer loci).

Possible eQTL relationships were explored using (i) the Genotype-Tissue Expression (GTEx v8) and (ii) the University of Barcelona and University of Virginia genotyping and RNA sequencing project (BarcUVa-Seq) dataset, which includes normal colon tissue samples from 445 healthy individuals [64]. In addition, the BarcUVa-Seq

project has data on physical activity in 352 (79%) participants, which we also used to test both specific eQTLs for physical activity status (active vs. inactive; study- and sex-specific quartile variable) and interactions between SNPs and physical activity on gene expression. The BarcUVa-Seq models were adjusted for age (years), sex, sequencing batch (one to four), and tissue location (left, right, transverse, missing). The putative functional role of SNPs and those in LD ($r^2 > 0.2$) and $MAF > 0.01$ at 500 kb flanking regions were investigated relative to their potential contribution to regulate gene expression by their physical location in regions of chromatin accessibility or histone modifications (variant enhancer loci). We annotated only suggestive eQTLs, i.e., those having a nominal p -value < 0.05 .

Details of the functional-annotations analyses have been previously published [65, 66]. Briefly, we used an assay for transposase-accessible chromatin with sequencing (ATAC-seq), DNaseI Hypersensitivity (DHS)-seq, H3K27ac histone ChIP-seq, and H3K4me1 histone ChIP-seq datasets of primary tissue from healthy colon and primary-tumor primary tissue samples containing active enhancer elements from Scacheri et al. [67], as well as from three CRC cell lines (SW480, HCT116, COLO205). These datasets were processed through ENCODE ATAC-seq/DNASE-seq [68] and histone ChIP-seq pipelines [69] to perform alignment and peak calling.

GxE analyses for rare variants

To assess the potential contribution of rare SNPs, we also performed gene-set-based aggregate tests only for rare SNPs using the Mixed effects Score Test for Interactions (MiSTi) approach [70] as a secondary analysis, as the power for rare SNPs testing usually is low. We examined the interactions of physical activity and aggregated rare SNP sets at the gene and enhancer level using MiSTi (MiSTi R package). We used a Fisher's combination approach under MiSTi (fMiSTi) to discover GxE interactions [70], after adjusting for age, sex, study, and the first three PCs. Because 25,000 gene regions were tested and this was a secondary analysis, interactions with $p < 2 \times 10^{-6}$ were considered statistically significant, whereas those with $p < 1 \times 10^{-4}$ were considered suggestive.

Results

Study population characteristics

The total sample size was $n = 39,992$ (16,383 CRC cases and 23,609 controls), with 76% classified as active (i.e., ≥ 8.75 MET-h/wk). Detailed descriptive characteristics of the study population are presented in Table 1. Compared to controls, CRC cases were more likely to be older, female, ever smokers, have a higher BMI and total

Table 1 Descriptive characteristics of all study participants by colorectal cancer case–control status with available physical activity data

Characteristics	Cases (N=16,383)	Controls (N=23,609)	P-value
Age (median imputed)^a			
Mean (SD)	65.0 (±9.4)	63.4 (±8.3)	<0.001
Sex			
Female	8677 (53%)	12,005 (51%)	<0.001
Male	7706 (47%)	11,604 (49%)	
Total energy consumption (kcal/day; mean imputed)^{b,c}			
Mean (SD)	1967 (±713)	1910 (±680)	<0.001
BMI (kg/m²)^c			
Mean (SD)	27.2 (±4.7)	26.9 (±4.5)	<0.001
Family history of colorectal cancer^c			
No	10,430 (64%)	12,945 (55%)	0.06
Yes	2295 (14%)	2685 (11%)	
Education level (highest completed)^c			
Less than high school	3070 (19%)	3488 (15%)	<0.001
High school/GED	3366 (21%)	3161 (13%)	
Some college	3476 (21%)	5783 (24%)	
College/graduate school	5601 (34%)	8488 (36%)	
Ever smoker^c			
No	7050 (43%)	11,479 (49%)	<0.001
Yes	9086 (55%)	11,862 (50%)	

Note: Data might not add to 100% because of rounding

Abbreviations: SD standard deviation, BMI body mass index, GED General Educational Development Test

Physical activity categorized as active (≥ 8.75 MET-h/wk) vs. inactive (< 8.75 MET-h/wk; reference category) dichotomous variable

^a Age was assessed at diagnosis or enrollment

^b Calculations exclude individuals with missing total energy intake information

^c Missing values not shown

P-values < 0.05 are statistically significant

energy consumption, and have a lower education level (each $p < 0.001$). Descriptive characteristics of the study population for the secondary physical activity variable assessed as study- and sex-specific quartiles are provided in Additional file 2: Table 2.

Physical activity and CRC risk

We observed that being active (≥ 8.75 MET-h/wk) vs. inactive (< 8.75 MET-h/wk) was associated with a 15% risk reduction in CRC in the overall meta-analysis (OR=0.85; 95% CI=0.81–0.90; Additional file 2: Fig. S1A; Additional file 2: Table 3). Sensitivity analyses showed even greater risk reduction for case–control studies (OR=0.75; 95% CI=0.66–0.85) compared to cohort-based studies (OR=0.88; 95% CI=0.83–0.93). No

evidence for heterogeneity was observed across all studies ($P_{\text{het}}=0.64$; $I^2=0\%$) or among case–control ($P_{\text{het}}=0.36$; $I^2=9\%$) or cohort-based studies ($P_{\text{het}}=0.91$; $I^2=0\%$). Further, analysis stratified by sex showed a risk reduction in both men (OR=0.83; 95% CI=0.76–0.90; $P_{\text{het}}=0.56$; $I^2=0\%$) and women (OR=0.87; 95% CI=0.81–0.94; $P_{\text{het}}=0.86$; $I^2=0\%$) when comparing active vs. inactive participants. For tumor site, the strongest inverse associations were observed for distal colon (OR=0.77; 95% CI=0.71–0.84; $P_{\text{het}}=0.64$; $I^2=0\%$) and proximal colon (OR=0.84; 95% CI=0.81–0.90; $P_{\text{het}}=0.46$; $I^2=0\%$), but not for rectal cancer (OR=0.94; 95% CI=0.85–1.04; $P_{\text{het}}=0.27$; $I^2=15\%$) comparing active vs. inactive participants. For physical activity measured as study- and sex-specific quartiles (treated as a continuous variable), we observed similar risk reductions for the overall meta-analysis as well as for stratified analysis by sex (Additional file 2: Fig. S1B; Additional file 2: Table 4). In dose–response (per-quartile) analyses, inverse associations were also observed for rectal cancer (per quartile OR=0.95; 95% CI=0.92–0.98; $P_{\text{het}} < 0.001$; $I^2=54\%$) as well as for distal and proximal colon, with some inter-study heterogeneity observed for case–control studies ($P_{\text{het}} < 0.001$; $I^2=74\%$). As we found statistically significant associations between physical activity and CRC for the overall population without significant evidence for heterogeneity, we conducted genome-wide GxE testing in the overall study population to maximize power.

Genome-wide physical activity–interaction scans for CRC risk

The Q-Q plot for the traditional gene–physical activity interactions for CRC risk using 1-d.f. analysis did not show p -value inflation for either primary or secondary physical activity variables (Additional file 2: Fig. S2).

Table 2 summarizes the statistically significant gene–physical activity interactions identified. Using the two-step EDGE method and the dichotomous physical activity variable (active vs. inactive), we identified statistically significant interactions for 5 SNPs, all of them in LD, on chromosome 15q13.3 located in the intergenic region between Gremlin 1 (*GREM1*) and Secretogranin V (*SCG5*) genes [71]. Among these SNPs with statistically significant interactions, we report only on the interaction of SNP rs4779584 with physical activity in this study (two-step p -value = 2.6×10^{-8} ; Table 2), as this SNP was supported by prior evidence on the association with CRC as a main effect (per T allele OR: active=1.20; 95% CI=1.10–1.20 vs. inactive=1.00; 95% CI=0.93–1.10; Table 3) [72]. This result remained robust in a sensitivity analysis that further accounted for BMI and interactions with BMI, as well as age, sex, study type, total energy consumption, and the first three PCs of genetic ancestry.

Table 2 Results of genome-wide interaction analyses with physical activity for colorectal cancer risk

Physical activity variable	SNP	Chr	BP position	Locus	Closest gene	Reference allele	Alternate allele	Alternate allele frequency	Type	Statistical method	P-value GxE^b
Active/inactive ^a	rs4779584	15	32,994,756	15q13.3	<i>GREM1</i> and <i>SCG5</i>	C	T	0.20	Intergenic variant	Two-step EDGE	2.6 × 10⁻⁸
Quartiles ^c	rs56906466	20	49,693,755	20q4.5	<i>KCNG1</i>	T	C	0.06	Intron	1-d.f. test	3.5 × 10⁻⁸

Abbreviations: *SNP* single nucleotide polymorphism, *Chr* chromosome, *BP position* base pair position based on NCBI Build 37, *1-d.f.* 1-degree of freedom

^a Physical activity categorized as active (≥ 8.75 MET-h/wk) vs. inactive (< 8.75 MET-h/wk; reference category)

^b *P*-value corresponds to the interaction between genetic variants (G) and physical activity (E) on risk of colorectal cancer in the combined case-control population based on the indicated statistical method

^c Physical activity assessed as study- and sex-specific quartiles

P-values that are statistically significant are indicated in bold text

Notes: Directly genotyped SNPs were coded as 0, 1, or 2 copies of the count allele. Imputed SNPs were coded as expected gene dosage. Multiplicative interaction terms were modelled as the product of PA and each SNP of interest

Table 3 Associations between physical activity for colorectal cancer risk stratified by genotypes of SNPs of interest

SNP	Physical activity	Homozygous non-carriers			Heterozygous			Homozygous carries of the alternate/minor allele			Per alternative allele within strata of physical activity categories	
		N (Ca/Co)	OR (95% CI)	P-value	N (Ca/Co)	OR (95% CI)	P-value	N (Ca/Co)	OR (95% CI)	P-value	OR (95% CI)	P-value
rs4779584	Inactive ^a	CC			CT			TT				
		2537/3642	1.00 (Ref.)	–	1304/1806	1.00 (0.95–1.10)	0.40	137/228	0.87 (0.69–1.10)	0.23	1.00 (0.93–1.10)	0.98
	Active ^a	7701/11,960	0.80 (0.75–0.85)	1.6 × 10⁻¹¹	4155/5372	0.95 (0.89–1.00)	0.19	549/601	1.10 (0.99–1.30)	0.08	1.20 (1.10–1.20)	2.0 × 10⁻¹⁵
	Active vs. inactive (by genotype)	–	0.80 (0.75–0.85)	1.6 × 10⁻¹¹	–	0.92 (0.84–1.00)	0.05	–	1.30 (1.00–1.70)	0.04		
rs56906466	Q1 ^b	TT			TC			CC				
		4168/5290	1.00 (Ref.)	–	443/715	0.77 (0.67–0.87)	8.0 × 10⁻⁵	20/19	1.10 (0.56–2.10)	0.81	0.77 (0.68–0.88)	1.4 × 10⁻⁴
	Q2 ^b	4085/5745	0.91 (0.85–0.96)	0.002	481/710	0.87 (0.76–0.99)	0.03	13/25	0.61 (0.3–1.20)	0.16	0.93 (0.82–1.10)	0.28
	Q3 ^b	3792/5896	0.81 (0.76–0.86)	6.8 × 10⁻¹²	469/669	0.87 (0.77–1.00)	0.047	21/26	1.00 (0.56–1.90)	0.96	1.10 (0.99–1.30)	0.08
	Q4 ^b	3342/5564	0.77 (0.72–0.82)	1.1 × 10⁻¹⁶	442/637	0.94 (0.82–1.10)	0.35	17/13	1.90 (0.91–4.20)	0.09	1.30 (1.10–1.50)	5.7 × 10⁻⁴
	Q2 vs. Q1 (by genotype)	–	0.91 (0.85–0.96)	0.002	–	1.10 (0.95–1.30)	0.16	–	0.56 (0.21–1.50)	0.23		
	Q3 vs. Q1 (by genotype)	–	0.81 (0.76–0.86)	6.8 × 10⁻¹²	–	1.10 (0.96–1.40)	0.14	–	0.93 (0.38–2.30)	0.88		
Q4 vs. Q1 (by genotype)	–	0.77 (0.72–0.82)	1.1 × 10⁻¹⁶	–	1.20 (1.00–1.50)	0.03	–	1.80 (0.65–4.90)	0.26			

Abbreviations: SNP single nucleotide polymorphism, PA physical activity, N number, Ca/Co case/control, OR odds ratio, 95% CI 95% confidence interval

Case/control counts were calculated by imputed genotype probabilities

^a Physical activity categorized as active (≥ 8.75 MET-h/wk) vs. inactive (< 8.75 MET-h/wk; reference category)

^b Physical activity, assessed as study- and sex-specific quartiles

P-values that are statistically significant are indicated in bold text

Specifically, these additional adjustments caused less than a 2% change in the G×Physical activity interaction estimates. Analysis stratified by rs4779584 genotype showed that participants who were physically active vs. inactive had 20% lower CRC risk among those who were carriers of CC (OR=0.80; 95% CI=0.75–0.85; $p=1.6 \times 10^{-11}$), while this risk reduction was diminished among those carrying the CT (OR=0.92; 95% CI=0.84–1.00) and TT (OR=1.30; 95% CI=1.00–1.70;) genotypes (Fig. 1; Table 3). We observed similar interaction effects when analyses were stratified for study type, sex, or tumor site (Additional file 2: Table 5).

The analysis of physical activity assessed as study- and sex-specific quartiles revealed an interaction with one SNP (rs56906466) on chromosome 20q4.5 located near the Potassium Voltage-Gated Channel Modifier Subfamily G Member 1 (*KCNKG1*) gene, using the traditional 1-d.f. test (G×E p -value= 3.5×10^{-8} ; Table 2; Additional file 2: Fig. S3B). This result was still consistent in a sensitivity analysis that also considered BMI and interactions with BMI along with age, sex, study type, total energy consumption, and the first three PCs of genetic ancestry. As in the previous sensitivity analysis, these adjustments resulted in less than a 2% variation in the G×Physical activity interaction estimates. Analysis stratified by rs56906466 genotype showed statistically significantly lower CRC risk with increases in physical activity, especially when comparing the highest quartile (Q4) to the lowest quartile (Q1), among those who were carriers of TT (OR=0.77; 95% CI=0.72–0.82; $p=1.1 \times 10^{-16}$). The corresponding inverse associations were not observed for those with TC (Q4 vs. Q1: OR=1.20; 95% CI=1.00–1.50; $p=0.03$) and CC (Q4 vs. Q1: OR=1.80; 95% CI=0.65–4.90; $p=0.26$) genotypes (Table 3). Similar interactions were observed when analyses were stratified by study type, sex, or tumor site (Additional file 2: Table 5). No other statistically significant interactions were observed (data not shown). Additionally, the G×E analyses for

rare variants did not identify any statistically significant interactions. There was also no significant LD-based correlation between rs4779584 and rs56906466 (correlation coefficient, $r^2=0.001$).

Functional follow-up

Functional annotation analyses around rs4779584 and rs56906466 showed enhanced activities. The SNP rs4779584 and correlated SNPs showed peaks in both normal (i.e., ATAC-seq, H3K4me1) and colon tumor samples (i.e., tumor DHS, tumor H3K27ac) as well as in cancer cell lines (i.e., H3K27ac, H3K4me1). The SNP rs56906466, although not correlated with other SNPs, was identified as a variant enhancer for tumor DHS and cell line DHS (Additional file 2: Figs. S4–5).

Two independent sources of eQTL analyses were used to expand on the regulatory roles of SNPs rs4779584 and rs56906466. The SNP rs4779584 was observed to be an eQTL in the GTEx v8 compendium as it modified the expression of *GREM1* in liver and pancreas, *SCG5* in liver, and RP11-758N13.1 in brain, cultured fibroblast, liver, and pancreas tissues. We did not observe any statistically significant eQTL findings for SNP rs56906466.

In relation to the BarcUVa-Seq dataset, which provides colon-specific eQTLs, the SNP in the 15q13.3 region did not modify the expression of *FNMI*, *GREM1*, *SCG5*, or other genes in the region (Additional file 2: Fig. S4). Likewise, the models tested in this dataset on the interaction with physical activity measured in the subjects did not reach statistical significance. The same approach was used to assess whether the SNP rs56906466 and the interaction term had eQTL effects on gene expression, but no statistically significant results were observed.

Discussion

To our knowledge, this is the largest genome-wide study conducted to date to investigate the interactions between variants across the genome and self-reported,

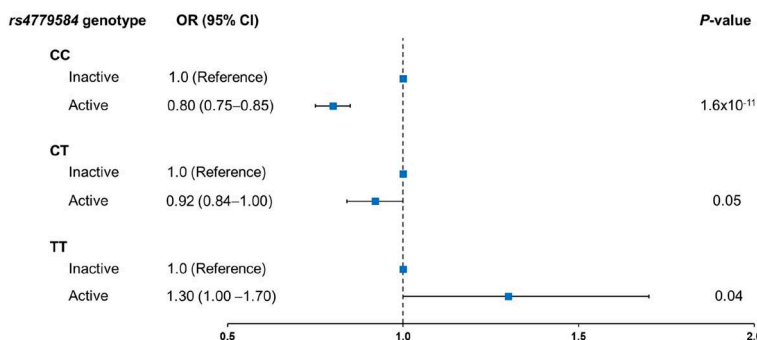


Fig. 1 Association between physical activity and colorectal cancer risk stratified by genotype of SNP rs4779584. Physical activity is categorized as active (≥ 8.75 MET-h/wk) vs. inactive (< 8.75 MET-h/wk; reference category)

harmonized physical activity data. Consistent with previous studies and the WCRF, we observed a statistically significant 15% lower CRC risk among physically active individuals, similar in magnitude to that previously observed [4, 10, 11, 73, 74]. Our analyses identified two novel, statistically significant GxE interactions for physical activity—SNPs rs4779584 and rs56906466 significantly modified the association between physical activity and CRC risk.

The SNP rs4779584, located in the 15q33.3 region, lies between the *GREM1* and *SCG5* genes and has been previously found to contribute to CRC susceptibility [28, 71, 75–78]. Carrying the T allele in rs4779584 has been reported to be associated with an increased CRC risk of 1.26 (95% CI=1.19–1.34) as compared to the C allele [79]. In our study, physical activity was significantly associated with a lower risk of CRC only among those with the C allele. Individuals with the CC genotype (reference group in our study, with the lowest baseline risk) derived the greatest benefit from physical activity, showing a statistically significant reduction in CRC risk among those who were physically active. Heterozygous CT individuals showed an intermediate level of risk and benefit from physical activity, consistent with an additive genetic model in which each additional T allele confers a progressive increase in CRC risk and attenuates the protective association of physical activity. Conversely, individuals with the TT genotype (the highest-risk group) did not appear to benefit from physical activity and instead showed a non-significant trend toward an increased risk. We hypothesize that these genotype-specific effects may reflect differential regulation of *GREM1* and/or *SCG5*-related signaling pathways, which could modulate the biological response to physical activity and influence colorectal carcinogenesis. *GREM1* encodes gremlin 1, which is a signaling protein involved in several pathways relevant to CRC, including the transforming growth factor- β (TGF- β) pathway, which has been implicated in tumor invasion and metastasis [80]. *GREM1* is also a proangiogenic factor, suggesting a possible role in cancer development when upregulated [81]. Additionally, Gremlin 1 is an insulin antagonist with elevated levels in type 2 diabetes [82], and has been linked to bone morphogenetic proteins (BMPs) signaling imbalance, which accelerates tumor cell proliferation [83], and is associated with inflammatory processes independently of BMPs [84, 85]. *SCG5* encodes secretogranin V (also named 7B2 protein or SGNE1), an essential neuroendocrine signaling molecule that plays a role in cellular proliferation [86, 87]. Although *SCG5* is associated with polyposis syndromes which are linked with CRC risk [88], its direct role in CRC is not as well characterized as compared to *GREM1*'s role in CRC [89]. Further, some studies have

also reported a role of *SCG5* in BMI modulation [90, 91]. The identified interactions suggest that the CRC risk reduction due to physical activity may be related to one or several more of these above-mentioned pathways.

There are only a small number of GWAS studies that have identified genetic loci associated with physical activity [92, 93], with one preclinical study suggesting that exercise training epigenetically reprograms *GREM1* expression [94]. However, to our knowledge, no prior studies have reported an interaction between rs4779584 and physical activity on CRC risk. Moreover, a recent genome-wide interaction study by Cho et al. using UK Biobank data with a modest sample size did not identify any statistically significant SNP–physical activity interactions, with all associations having FDR-adjusted p -values >0.05 [26]. Although gene-level analysis yielded signals for *GREM1* and *SCG5* (comprising 1 and 26 SNPs, respectively), both genes had adjusted p -values >0.98 for the gene–physical activity interaction, and no additional details were provided. The epidemiologic evidence indicating the beneficial effect of physical activity on CRC risk is extensive, and several biological mechanisms have been identified or proposed, including in some intervention studies, such as physical activity's effect on immune system, systemic inflammatory markers, energy regulation, hormones levels, insulin resistance, and gut microbial composition [5, 7, 95, 96]. Related to our findings, a randomized trial conducted in obese patients who followed different resistance training protocols observed significant reductions in plasma gremlin 1 and C-reactive protein levels compared to a control group [97]. Additionally, myokines (i.e., cytokines), such as myostatin (member of the TGF- β family) or interleukin-6, are secreted by the skeletal muscle in response to intensity training [98, 99]. The effect of regular exercise on *SCG5*, the other gene close to the SNP rs4779584 that showed interactions with physical activity on CRC risk, has been investigated in experimental studies using animal models. However, the results were inconclusive, with one study reported non-significantly decreased *SCG5* expression, while the other study reported significantly increased expression levels [100, 101]. Future studies are warranted to describe the plausible biological mechanism by which SNP rs4779584 interacts with physical activity and modifies CRC risk, but on the basis of our findings, genetic markers in this region showed enhanced activity in both normal and tumor samples suggesting a potential regulatory role on transcription of adjacent genes. Consistent with this, we observed that SNP rs4779584 modified the expression of *GREM1* and *SCG5* in pancreas and liver, but not in colon tissue. These results were based solely on GTEx data, as no significant associations were observed in the BarcUva-Seq dataset, which is specific to colon

tissue. Due to its small sample size, statistical power in BarcUVa-Seq was likely insufficient to detect modest gene expression effects or eQTL \times physical activity interactions; therefore, these findings should be considered exploratory.

We also discovered a new locus rs56906466 located near *KCNQ1* that has not been previously associated with CRC, physical activity, or its interaction with physical activity on CRC risk. Notably, Cho et al., using the UK Biobank data, also reported three SNPs within the broader *KCNQ1* region; however, the specific variants were not disclosed, and the corresponding adjusted p-value for the gene–physical activity interaction was 0.99, indicating no evidence of a statistically significant association [26]. This gene encodes a member of the large gene family that instructs the building of potassium channels and is abundantly expressed in skeletal muscle. *KCNQ1* has been related to insulin secretion, muscle contraction, and neurotransmitter release regulation, among others [102]; however, its functions are not fully understood. Our findings showed that rs56906466 had statistically significant interactions with physical activity in modifying CRC risk. Furthermore, functional-annotations analyses demonstrated that some of the genetic variants interacting with physical activity were located in enhancers and were linked to differential gene expression. However, additional targeted studies will be necessary to further investigate the joint effects of these genes with physical activity on CRC risk.

There is increasing evidence that gene–physical activity interactions (including being physically active or inactive) have an effect on several health-related outcomes such as blood pressure, hypertension, BMI, and insulin metabolism [103]. However, few studies have evaluated the gene–physical activity interaction on CRC risk, and all previous studies followed a candidate-gene approach and included only a limited number of SNPs [22–25]. Two studies evaluated the mediating effects of physical activity on CRC risk via alterations in polymorphisms in the insulin-like growth factor-1 (*IGF-1*) gene, since physical activity is known to modulate IGF-1 serum levels, and observed statistically significant interactions [22, 104]. Khoury-Shakour et al. focused their analysis on the polymorphism rs2665802 at intron 4 of the growth hormone 1 (*GHI*) gene and observed that the minor allele A was associated with lower risk of CRC among inactive participants [22]. Another study assessed the interaction between physical activity and CRC risk based on a polymorphism (rs647161) in the paired-like homeodomain 1 (*PITX1*) gene in a Korean population, and reported a higher risk of CRC among participants who exercised less and carried the minor allele [23]. *PITX1* is considered a tumor suppressor gene [105], and is known to

influence the expression of *GHI*, and is related to *IGF-1* [106]. Song et al. assessed interactions between physical activity and 31 SNPs (including rs4779584) on CRC risk among 703 CRC cases and 1406 healthy controls [24]. However, they observed statistically significant interactions only with rs4444235—with increased CRC risk among C carriers who exercised regularly—but not for rs4779584, which may be due to the small sample size. However, none of the above findings could be replicated in the present study (data not shown). Additionally, we observed no LD-based correlation between rs4779584 and rs4444235 ($r^2=0.004$). Given the smaller sample size and candidate gene approach in the study by Song et al., it is possible that these are chance findings.

A main strength of our study was a large, well-characterized study population, the largest to date to have examined gene–physical activity interactions. The use of several complementary statistical approaches was also a strength of this study as it allowed detection of specific loci within *GREM1* and *SCG5* and near *KCNQ1* genes. However, our findings may not be generalizable outside of European-descent populations, as the participants in this study were limited to those with European ancestry and were, on average, more physically active than the US general population. Although models were adjusted for the first three genetic principal components to account for subtle population structure, some residual confounding cannot be excluded. The consortium is actively striving to overcome this limitation by expanding our research to include other racial and ethnic groups, as well as by harmonizing epidemiological data, which will enable us to expand our future G \times E analyses. Additionally, this study included self-report measures of physical activity, which are prone to recall and response biases, but these are likely to attenuate “true” associations with disease risk [107]. Although physical activity was harmonized across studies as MET-h/wk using established methods, differences in questionnaire design, recall periods, and assessment timing, as well as limited detail on activity types, may have introduced potential exposure misclassification. In addition, the time interval between physical activity assessment and CRC diagnosis varied across studies, which may have affected exposure accuracy. In some cases, such as in case–control studies, physical activity levels may have been influenced by undiagnosed symptoms or preclinical disease, potentially introducing reverse causation bias; however, it is important to note that over 70% of the studies included were cohort studies. These limitations could have reduced our ability to detect more subtle G \times E interactions involving physical activity. Moreover, as the study is based on observational and genetic data, its design does not allow for causal inferences. Our findings have also not yet been

replicated in an independent dataset, which may increase the likelihood of false-positive associations and should be considered when interpreting the results. Lastly, our sample size did not allow us to identify genes whose rare variants may interact with physical activity in aggregate tests, and further functional studies are needed to verify the role of the identified SNPs in modulating CRC risk through physical activity.

Conclusions

In conclusion, we identified two novel genetic loci that interact with physical activity and modify the association between physical activity and CRC risk. Potential mechanisms behind the interaction of rs4779584 and physical activity in CRC risk may be linked in part to BMP-related, inflammation, and/or insulin signaling pathways in response to physical activity. However, SNP rs56906466, which is near a potassium channel gene, has not been previously described in relation to physical activity or CRC, and additional investigations are required to elucidate the potential mechanisms through which it may be involved in colorectal carcinogenesis, especially in individuals who are not physically active.

Abbreviations

CRC	Colorectal cancer
WCRF/AICR	World Cancer Research Fund/American Institute for Cancer Research
GxE	Gene-environment
CCFR	Colon Cancer Family Registry
GECCO	Genetics and Epidemiology of Colorectal Cancer Consortium
CORECT	Colorectal Cancer Transdisciplinary Study
CDEs	Common data elements
BMI	Body mass index
IPAQ	International Physical Activity Questionnaire
EPIC	European Prospective Investigation into Cancer and Nutrition
MET-h/wk	Metabolic equivalent task hours per week
SNPs	Single nucleotide polymorphisms
HRC	Haplotype Reference Consortium
MAF	Minor allele frequency
LD	Linkage disequilibrium
PCs	Principal components
ORs	Odds ratios
CI	Confidence intervals
EDGE	Two-step screening and testing method
PCA	Principal component analysis
Q-Q	Quantile-quantile
eQTLs	Expression quantitative trait loci
GTEx v8	Genotype-Tissue Expression
BarCUVa-Seq	University of Barcelona and University of Virginia genotyping and RNA sequencing project
ATAC-seq	Assay for transposase-accessible chromatin with sequencing
DHS	DNaseI Hypersensitivity
MiSTi	Mixed effects Score Test for Interactions
GREM1	Gremlin 1
SCG5	Secretogranin V
KCNG1	Potassium Voltage-Gated Channel Modifier Subfamily G Member 1
TGF-β	Transforming growth factor-β
BMPs	Bone morphogenetic proteins
IGF-1	Insulin-like growth factor-1
GH1	Growth hormone 1
PITX1	Paired-like homeodomain 1

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-026-04675-5>.

Additional file 1: Table 1. Description and count of participants per study included in the analysis.

Additional file 2: Tables 2–5; Figures S1–S5. Table 2: Descriptive characteristics of all study participants by case-control status, who had study- and sex-specific quartile physical activity data available. Table 3: Associations between physical activity and colorectal cancer risk stratified by study design, sex, and tumor site. Table 4: Association between physical activity and colorectal cancer risk stratified by study design. Table 5: Associations between the identified SNPs and physical activity for colorectal cancer risk stratified by study design, sex, and tumor site. Figure S1. Forest plot showing the association between physical activity and colorectal cancer risk for studies included in the gene-physical activity interaction analysis, adjusted for age at diagnosis or enrollment, sex, and total energy consumption (kcal/day, when available). A: Dichotomized physical activity (i.e., active: ≥ 8.75 MET-h/wk vs. inactive: < 8.75 MET-h/wk). B: Physical activity assessed as study- and sex-specific quartiles (treated as continuous variable). Figure S2. Quantile-Quantile (Q-Q) plots for the traditional 1-d.f. interaction tests of physical activity for colorectal cancer risk. A: Dichotomized physical activity (i.e., active: ≥ 8.75 MET-h/wk vs. inactive: < 8.75 MET-h/wk). B: Physical activity assessed as study- and sex-specific quartiles (treated as continuous variable). Figure S3. Manhattan plots for the traditional 1-d.f. multiplicative interaction tests of SNPs with physical activity for colorectal cancer risk, adjusted for age, sex, study, total energy consumption (kcal/day, when available), and the first three principal components. A: Dichotomized physical activity (i.e., active: ≥ 8.75 MET-h/wk vs. inactive: < 8.75 MET-h/wk). B: Physical activity assessed as study- and sex-specific quartiles (treated as continuous variable). Figure S4. Regional (LocusZoom) plots for interaction result of SNPs x physical activity for colorectal cancer risk. A: Dichotomized physical activity (i.e., active: ≥ 8.75 MET-h/wk vs. inactive: < 8.75 MET-h/wk). B: Physical activity assessed as study- and sex-specific quartiles (treated as continuous variable). Figure S5. Functional annotation plots showing chromatin accessibility for genetic regions that showed significant interaction with physical activity for colorectal cancer risk. A: rs4779584/GREM1 region. B: rs56906466/KCNG1 region. Top panel indicates GENCODE reference genes (GRCh37).

Acknowledgements

CCFR: The Colon CFR graciously thanks the generous contributions of their study participants, dedication of study staff, and the financial support from the U.S. National Cancer Institute, without which this important registry would not exist. The authors would like to thank the study participants and staff of the Seattle Colon Cancer Family Registry and the Hormones and Colon Cancer study (CORE Studies).

CPS-II: The authors express sincere appreciation to all Cancer Prevention Study-II participants, and to each member of the study and biospecimen management group. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries and cancer registries supported by the National Cancer Institute's Surveillance Epidemiology and End Results Program. The study protocol was approved by the institutional review boards of Emory University, and those of participating registries as required. The authors assume full responsibility for all analyses and interpretation of results. The views expressed here are those of the authors and do not necessarily represent the American Cancer Society or the American Cancer Society – Cancer Action Network.

DACHS: We thank all participants and cooperating clinicians, and everyone who provided excellent technical assistance.

EPIC: Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/World Health Organization.

Harvard cohorts (HPFS, NHS): The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H.

Chan School of Public Health, and those of participating registries as required. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. Central registries may also be supported by state agencies, universities, and cancer centers. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming.

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: <https://s3-us-west-2.amazonaws.com/www-who-org/wp-content/uploads/WHI-Investigator-Long-List.pdf>

Disclaimer

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policies, or views of the International Agency for Research on Cancer/World Health Organization. This article is the result of the scientific work of Dr. Murphy, Dr. Papadimitriou, and Dr. Dimou while they were affiliated at IARC.

Authors' contributions

All authors participated in the revisions to this paper and the interpretation of the results, and all authors read and approved the final paper. *Conceptualization*: A.R.P., M.O.-S., W.J.G., U.P., V.M., C.M.U., A.E.K., E.S.K., L.M., Y.L. *Data curation*: A.E.K., E.S.K., C.Q., F.M.-N., J.M., Y.L. *Formal analysis*: Q.F., A.E.K., E.S.K., C.Q., F.M.-N., J.M., Y.L. *Methodology*: A.R.P., M.O.-S., W.J.G., U.P., V.M., A.E.K., E.S.K., C.Q., F.M.-N., J.M., Y.L. *Writing—original draft*: A.R.P., M.O.-S., W.J.G., U.P., V.M., C.M.U., A.E.K., E.S.K., J.M., Y.L. *Writing—Review and editing*: All authors. *Supervision*: U.P., W.J.G., C.M.U., V.M.

Funding

Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088, R01 CA059045, U01 CA164930, R21 CA191312, R01201407, R01CA488857, R01CA273198, R01CA244588). Genotyping/Sequencing services were provided by the Center for Inherited Disease Research (CIDR) contract number HHSN2682017000061 and HHSN2682012000081. This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704. Scientific Computing Infrastructure at Fred Hutch funded by ORIP grant S10OD028685. Statistical methodology and software development at USC funded by P01CA196569.

Colon Cancer Family Registry (CCFR): CCFR (www.coloncf.org) is supported in part by funding from the National Cancer Institute (NCI), National Institutes of Health (NIH) (award U01 CA167551). Support for case ascertainment was provided in part from the Surveillance, Epidemiology, and End Results (SEER) Program and the following U.S. state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada). The CCFR Set-1 (Illumina 1 M/1 M-Duo) and Set-2 (Illumina Omni1-Quad) scans were supported by NIH awards U01 CA122839 and R01 CA143247 (to GC). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and R01 CA81488 (to SBG). The CCFR Set-4 (Illumina OncoArray 600 K SNP array) was supported by NIH award U19 CA148107 (to SBG) and by the Center for Inherited Disease Research (CIDR), which is funded by the NIH to the Johns Hopkins University, contract number HHSN2682012000081. Additional funding for the OFCCR/ARCTIC was through award GL201-043 from the Ontario Research Fund (to BWZ), award 112746 from the Canadian Institutes of Health Research (to TJH), through a Cancer Risk Evaluation (CaRE) Program grant from the Canadian Cancer Society (to SG), and through generous support from the Ontario Ministry of Research and Innovation. The SFCCR Illumina HumanCytoSNP array was supported in part through NCI/NIH awards U01/U24 CA074794 and R01 CA076366 (to PAN). The content of this manuscript does not necessarily reflect the views or policies of the NCI, NIH or any of the collaborating centers in the Colon Cancer Family

Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the CCFR.

COLO2&3: National Institutes of Health (R01 CA060987).

Colorectal Cancer Transdisciplinary (CORECT) Study: The CORECT Study was supported by the National Cancer Institute, National Institutes of Health (NCI/NIH), U.S. Department of Health and Human Services (grant numbers U19 CA148107, R01 CA81488, P30 CA014089, R01 CA197350; P01 CA196569; R01 CA201407) and National Institutes of Environmental Health Sciences, National Institutes of Health (grant number T32 ES013678).

CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (CPS-II) cohort. The study protocol was approved by the institutional review boards of Emory University and those of participating registries as required.

DACHS: This work was supported by the German Research Council (BR 1704/6–1, BR 1704/6–3, BR 1704/6–4, CH 117/1–1, HO 5117/2–1, HE 5998/2–1, KL 2354/3–1, RO 2270/8–1 and BR 1704/17–1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A and 01ER1505B).

DALS: National Institutes of Health (R01 CA48998 to M. L. Slattery).

EDRN: This work is funded and supported by the NCI, EDRN Grant (U01 CA 84968–06).

EPIC: The coordination of EPIC is financially supported by the International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Éducation Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam-Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS)—Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology—ICO (Spain); Swedish Cancer Society, Swedish Research Council and County Councils of Skåne and Västerbotten (Sweden); Cancer Research UK (14,136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1,000,143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford) (United Kingdom).

Harvard cohorts (HPFS, NHS): HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), and NHS is supported by the National Institutes of Health (R01 CA137178, P01 CA087969, UM1 CA186107, R01 CA151993, and R35 CA197735).

Hawaii Adenoma Study: NCI grants R01 CA072520.

LCCS: The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167).

MEC: National Institutes of Health (R37 CA054281, P01 CA033619, and R01 CA063464).

NCCCS I & II: We acknowledge funding support for this project from the National Institutes of Health, R01 CA66635 and P30 DK034987.

NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821); the National Institutes of Health, U.S. Department of Health and Human Services (U01 CA74783); and National Cancer Institute of Canada grants (18,223 and 18,226). The authors wish to acknowledge the contribution of Alexandre Belisle and the genotyping team of the McGill University and Génomique Québec Innovation Centre, Montréal, Canada, for genotyping the Sequenom panel in the NFCCR samples. Funding was provided to Michael O. Woods by the Canadian Cancer Society Research Institute.

Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the Swedish Research Council/Infrastructure grant, the Swedish Cancer Foundation, and the Karolinska Institute's Distinguished Professor Award to Alicja Wolk.

UK Biobank: This research has been conducted using the UK Biobank Resource under Application Number 8614.

VITAL: National Institutes of Health (K05 CA154337).

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts 75N92021D00001, 75N92021D00002, 75N92021D00003, 75N92021D00004, 75N92021D00005.

Data availability

The dataset used in the current study may be available from the corresponding author on reasonable request for researchers who meet the criteria for access to confidential data.

Declarations

Ethics approval and consent to participate

The study was conducted in accordance with the principles of the Declaration of Helsinki; each contributing study was approved by its respective Institutional Review Board or relevant research committee. For CPS-II, written informed consent was received from participants to obtain medical records. At the time of each mailed survey, participants were informed that their identifying information would be used to link with cancer registries and death indexes. For the other studies, all study participants provided informed consent.

Consent for publication

Not applicable.

Competing interests

Dr. Ulrich has as HCl Cancer Center Director oversight over research funded by several pharmaceutical companies but has not received funding directly herself. Dr. Peters was a consultant with AbbVie and her husband is holding individual stocks for the following companies: BioNTech SE – ADR, Amazon, CureVac BV, NanoString Technologies, Google/Alphabet Inc Class C, NVIDIA Corp, Microsoft Corp. Other authors declare that they have no conflict of interest.

Author details

¹Huntsman Cancer Institute, Salt Lake City, UT, USA. ²Department of Population Health Sciences, University of Utah, Salt Lake City, UT, USA. ³Department of Population Science, American Cancer Society, Atlanta, GA, USA. ⁴Unit of Biomarkers and Susceptibility (UBS), Oncology Data Analytics Program (ODAP), Catalan Institute of Oncology (ICO), L'Hospitalet del Llobregat, Barcelona 08908, Spain. ⁵ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona 08908, Spain. ⁶Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid 28029, Spain. ⁷Division of Biostatistics, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. ⁸Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA, USA. ⁹Department of Clinical Sciences, Faculty of Medicine and Health Sciences and Universitat de Barcelona Institute of Complex Systems (UBICS), University of Barcelona (UB), L'Hospitalet de Llobregat, Barcelona 08908, Spain. ¹⁰Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA. ¹²Leeds Institute of Cancer and Pathology, University of Leeds, Leeds, UK. ¹³Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece. ¹⁴Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany. ¹⁵German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. ¹⁶Colorectal Oncogenomics Group, Department of Clinical Pathology, The University of Melbourne, Parkville, VIC 3010, Australia. ¹⁷University of Melbourne Centre for Cancer Research, Victorian Comprehensive Cancer Centre, Parkville, VIC 3010, Australia. ¹⁸Genomic Medicine and Family Cancer Clinic, The Royal Melbourne Hospital, Parkville, VIC, Australia. ¹⁹Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY, USA. ²⁰Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. ²¹Channing Division of Network Medicine, Brigham

and Women's Hospital and Harvard Medical School, Boston, MA, USA. ²²Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. ²³Broad Institute of MIT and Harvard, Cambridge, MA, USA. ²⁴Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA. ²⁵Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA. ²⁶Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany. ²⁷University Medical Centre Hamburg-Eppendorf, University Cancer Centre Hamburg (UCCH), Hamburg, Germany. ²⁸Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. ²⁹Department of Gastroenterology, Kaiser Permanente Medical Center, San Francisco, CA, USA. ³⁰Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. ³¹Department of Public Health Sciences, Center for Public Health Genomics, Charlottesville, VA, USA. ³²Nutrition and Metabolism Branch, International Agency for Research On Cancer, World Health Organization, Lyon, France. ³³Department of Medical Oncology & Therapeutics Research, City of Hope National Medical Center, Duarte, CA, USA. ³⁴Center for Precision Medicine, City of Hope National Medical Center, Duarte, CA, USA. ³⁵Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. ³⁶Department of Radiation Sciences, Oncology Unit, Umeå University, Umeå, Sweden. ³⁷Department of Biostatistics, University of Washington, Seattle, WA, USA. ³⁸Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, NC, USA. ³⁹Department of Genetics, Stanford University, Stanford, CA, USA. ⁴⁰Department of Computer Science, Stanford University, Stanford, CA, USA. ⁴¹Department of Family Medicine, University of Virginia, Charlottesville, VA, USA. ⁴²UVA Comprehensive Cancer Center, University of Virginia, Charlottesville, VA, USA. ⁴³Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, VIC, Australia. ⁴⁴Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia. ⁴⁵University of Hawaii Cancer Center, Honolulu, HI, USA. ⁴⁶The Research Group in Gene - Environment and Health Interactions (GILGAS), Institut of Biomedicine (IBIOMED), Universidad de León, León 24071, Spain. ⁴⁷Faculty of Health Sciences, Department of Biomedical Sciences, Area of Preventive Medicine and Public Health, Universidad de León, León 24071, Spain. ⁴⁸Program in MPE Molecular Pathological Epidemiology, Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. ⁴⁹Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA. ⁵⁰Department of Media, Information, and Design, University of Applied Sciences and Arts, Hannover, Germany. ⁵¹Department of Laboratory Medicine and Pathology, Mayo Clinic Arizona, Scottsdale, AZ, USA. ⁵²Slone Epidemiology Center at Boston University, Boston, MA, USA. ⁵³Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia. ⁵⁴Department of Cancer Medicine, University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁵⁵Department of Community Medicine and Epidemiology, Lady Davis Carmel Medical Center, Haifa, Israel. ⁵⁶Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ⁵⁷Research Centre for Hauora and Health, Massey University, Wellington, New Zealand. ⁵⁸Department of Epidemiology, University of Washington School of Public Health, Seattle, WA, USA. ⁵⁹Technion-Israel Institute of Technology and APRPM, Haifa, Israel. ⁶⁰Clalit National Cancer Control Center, Haifa, Israel. ⁶¹Department of Nutritional Sciences, University of Michigan School of Public Health, Ann Arbor, MI, USA. ⁶²Departments of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA. ⁶³Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. ⁶⁴School of Public Health, Capital Medical University, Beijing, China. ⁶⁵Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, The Netherlands. ⁶⁶Wallenberg Centre for Molecular Medicine, Umeå University, Umeå, Sweden. ⁶⁷Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, USA. ⁶⁸Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. ⁶⁹Division of Biomedical Sciences, Memorial University of Newfoundland, St. John's, Canada. ⁷⁰Biostatistics Division, Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA.

Received: 12 August 2025 Accepted: 29 January 2026
Published online: 05 February 2026

References

- Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2024;74(3):229–63.
- Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017;66(4):683–91.
- Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. *Transl Oncol*. 2021;14(10):101174.
- World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Expert Report 2018. Diet, nutrition, physical activity and colorectal cancer. Available at dietandcancerreport.org. In.
- Ulrich CM, Himmert C, Holowatyj AN, Hursting SD. Energy balance and gastrointestinal cancer: risk, interventions, outcomes and mechanisms. *Nat Rev Gastroenterol Hepatol*. 2018;15(11):683–98.
- Amirsasan R, Akbarzadeh M, Akbarzadeh S. Exercise and colorectal cancer: prevention and molecular mechanisms. *Cancer Cell Int*. 2022;22(1):247.
- Jurdana M. Physical activity and cancer risk. Actual knowledge and possible biological mechanisms. *Radiol Oncol*. 2021;55(1):7–17.
- Stein MJ, Baurecht H, Bohmann P, Fervers B, Fontvieille E, Freisling H, et al. Diurnal timing of physical activity and risk of colorectal cancer in the UK Biobank. *BMC Med*. 2024;22(1):399.
- Papadimitriou N, Dimou N, Tsilidis KK, Banbury B, Martin RM, Lewis SJ, et al. Physical activity and risks of breast and colorectal cancer: a Mendelian randomisation analysis. *Nat Commun*. 2020;11(1):597.
- McTiernan A, Friedenreich CM, Katzmarzyk PT, Powell KE, Macko R, Buchner D, et al. Physical activity in cancer prevention and survival: a systematic review. *Med Sci Sports Exerc*. 2019;51(6):1252–61.
- Matthews CE, Moore SC, Arem H, Cook MB, Trabert B, Hakansson N, et al. Amount and intensity of leisure-time physical activity and lower cancer risk. *J Clin Oncol*. 2020;38(7):686–97.
- Strain T, Flaxman S, Guthold R, Semenova E, Cowan M, Riley LM, et al. National, regional, and global trends in insufficient physical activity among adults from 2000 to 2022: a pooled analysis of 507 population-based surveys with 5.7 million participants. *Lancet Glob Health*. 2024;12(8):e1232–43.
- Fiuzza-Luces C, Valenzuela PL, Galvez BG, Ramirez M, Lopez-Soto A, Simpson RJ, et al. The effect of physical exercise on anticancer immunity. *Nat Rev Immunol*. 2024;24(4):282–93.
- Xie W, Lu D, Liu S, Li J, Li R. The optimal exercise intervention for sleep quality in adults: a systematic review and network meta-analysis. *Prev Med*. 2024;183:107955.
- Noetel M, Sanders T, Gallardo-Gomez D, Taylor P, Del Pozo Cruz B, van den Hoek D, et al. Effect of exercise for depression: systematic review and network meta-analysis of randomised controlled trials. *BMJ*. 2024;384:e075847.
- Chalitsios CV, Markozannes G, Aglago EK, Berndt SI, Buchanan DD, Campbell PT, Cao Y, Chan AT, Dimou N, Drew DA et al. Physical activity and molecular subtypes of colorectal cancer: a pooled observational analysis and Mendelian randomization study. *JNCI Cancer Spectr*. 2025;9(6):pkaf095. <https://pubmed.ncbi.nlm.nih.gov/41031512/>.
- Bouras E, Yu R, Kim AE, Markozannes G, Murphy N, Albanes D, et al. Using gene-environment interactions to explore pathways for colorectal cancer risk. *EBioMedicine*. 2025;121:105964.
- Ungvari Z, Fekete M, Varga P, Lehoczki A, Fekete JT, Ungvari A, et al. Overweight and obesity significantly increase colorectal cancer risk: a meta-analysis of 66 studies revealing a 25–57% elevation in risk. *Geroscience*. 2025;47(3):3343–64.
- Li W, Liu T, Qian L, Wang Y, Ma X, Cao L, et al. Insulin resistance and inflammation mediate the association of abdominal obesity with colorectal cancer risk. *Front Endocrinol (Lausanne)*. 2022;13:983160.
- Urban S, Chmura O, Wator J, Panek P, Zapala B. The intensive physical activity causes changes in the composition of gut and oral microbiota. *Sci Rep*. 2024;14(1):20858.
- Himmert C, Stephens WZ, Gigic B, Hardikar S, Holowatyj AN, Lin T, et al. Differences in the gut microbiome by physical activity and BMI among colorectal cancer patients. *Am J Cancer Res*. 2022;12(10):4789–801.
- Khoury-Shakour S, Gruber SB, Lejbkovicz F, Rennert HS, Raskin L, Pinchev M, et al. Recreational physical activity modifies the association between a common GH1 polymorphism and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2008;17(12):3314–8.
- Gunathilake MN, Lee J, Cho YA, Oh JH, Chang HJ, Sohn DK, et al. Interaction between physical activity, PITX1 rs647161 genetic polymorphism and colorectal cancer risk in a Korean population: a case-control study. *Oncotarget*. 2018;9(7):7590–603.
- Song N, Lee J, Cho S, Kim J, Oh JH, Shin A. Evaluation of gene-environment interactions for colorectal cancer susceptibility loci using case-only and case-control designs. *BMC Cancer*. 2019;19(1):1231.
- Morimoto LM, Newcomb PA, White E, Bigler J, Potter JD. Insulin-like growth factor polymorphisms and colorectal cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2005;14(5):1204–11.
- Cho S, Shin A. Genome-wide interaction study of physical activity and genetic susceptibility on colorectal cancer using UK biobank data. *Sci Rep*. 2025;15(1):30180.
- Huyghe JR, Bien SA, Harrison TA, Kang HM, Chen S, Schmit SL, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet*. 2019;51(1):76–87.
- Peters U, Jiao S, Schumacher FR, Hutter CM, Aragaki AK, Baron JA, Berndt SI, Bezieau S, Brenner H, Butterbach K et al. Identification of Genetic Susceptibility Loci for Colorectal Tumors in a Genome-Wide Meta-analysis. *Gastroenterology* 2013;144(4):799–807 e724.
- Schmit SL, Edlund CK, Schumacher FR, Gong J, Harrison TA, Huyghe JR, et al. Novel Common Genetic Susceptibility Loci for Colorectal Cancer. *J Natl Cancer Inst*. 2019;111(2):146–57.
- Schumacher FR, Schmit SL, Jiao S, Edlund CK, Wang H, Zhang B, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat Commun*. 2015;6:7138.
- Hutter CM, Chang-Claude J, Slattery ML, Pflugeisen BM, Lin Y, Duggan D, et al. Characterization of gene-environment interactions for colorectal cancer susceptibility loci. *Cancer Res*. 2012;72(8):2036–44.
- Wang X, O'Connell K, Jeon J, Song M, Hunter D, Hoffmeister M, et al. Combined effect of modifiable and non-modifiable risk factors for colorectal cancer risk in a pooled analysis of 11 population-based studies. *BMJ Open Gastroenterol*. 2019;6(1):e000339.
- Rolland B, Reid S, Stelling D, Warnick G, Thornquist M, Feng Z, et al. Toward rigorous data harmonization in cancer epidemiology research: one approach. *Am J Epidemiol*. 2015;182(12):1033–8.
- Littman AJ, White E, Kristal AR, Patterson RE, Satia-Abouta J, Potter JD. Assessment of a one-page questionnaire on long-term recreational physical activity. *Epidemiology*. 2004;15(1):105–13.
- Nelson ME, Rejeski WJ, Blair SN, Duncan PW, Judge JO, King AC, et al. Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. *Med Sci Sports Exerc*. 2007;39(8):1435–45.
- Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The Physical Activity Guidelines for Americans. *JAMA*. 2018;320(19):2020–8.
- Schmitz KH, Courneya KS, Matthews C, Demark-Wahnefried W, Galvao DA, Pinto BM, et al. American College of Sports Medicine roundtable on exercise guidelines for cancer survivors. *Med Sci Sports Exerc*. 2010;42(7):1409–26.
- Kushi LH, Doyle C, McCullough M, Rock CL, Demark-Wahnefried W, Bandera EV, et al. American Cancer Society guidelines on nutrition and physical activity for cancer prevention: reducing the risk of cancer with healthy food choices and physical activity. *CA Cancer J Clin*. 2012;62(1):30–67.
- WHO guidelines on physical activity and sedentary behaviour. Geneva: World Health Organization; 2020. Licence: CC BY-NC-SA 3.0 IGO.
- Phipps AI, Shi Q, Zemla TJ, Dotan E, Gill S, Goldberg RM, et al. Physical activity and outcomes in patients with stage III colon cancer: a correlative analysis of phase III trial NCCTG N0147 (Alliance). *Cancer Epidemiol Biomarkers Prev*. 2018;27(6):696–703.
- Hardikar S, Newcomb PA, Campbell PT, Win AK, Lindor NM, Buchanan DD, et al. Prediagnostic physical activity and colorectal cancer survival: overall and stratified by tumor characteristics. *Cancer Epidemiol Biomarkers Prev*. 2015;24(7):1130–7.
- Kuiper JG, Phipps AI, Neuhaus ML, Chlebowski RT, Thomson CA, Irwin ML, et al. Recreational physical activity, body mass index, and

- survival in women with colorectal cancer. *Cancer Causes Control*. 2012;23(12):1939–48.
43. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48(10):1284–7.
 44. Morrison J. <https://cran.r-project.org/package=BinaryDosage>. 2020.
 45. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56–65.
 46. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol*. 2014;14:25.
 47. Cochran WG. The combination of estimates from different experiments. *Biometrics*. 1954;10:101–29.
 48. Guido Schwarzer JRC, Gerta Rücker. *Meta-Analysis with R*. Switzerland: Springer, Cham; 2015.
 49. Dai JY, Logsdon BA, Huang Y, Hsu L, Reiner AP, Prentice RL, et al. Simultaneously testing for marginal genetic association and gene-environment interaction. *Am J Epidemiol*. 2012;176(2):164–73.
 50. Gauderman WJ, Kim A, Conti DV, Morrison J, Thomas DC, Vora H, et al. A unified model for the analysis of gene-environment interaction. *Am J Epidemiol*. 2019;188(4):760–7.
 51. Gao X, Starmer J, Martin ER. A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet Epidemiol*. 2008;32(4):361–9.
 52. Kawaguchi ES, Kim AE, Lewinger JP, Gauderman WJ. Improved two-step testing of genome-wide gene-environment interactions. *Genet Epidemiol*. 2023;47(2):152–66. <https://pubmed.ncbi.nlm.nih.gov/36571162/>.
 53. Gauderman WJ, Zhang P, Morrison JL, Lewinger JP. Finding novel genes by testing $G \times E$ interactions in a genome-wide association study. *Genet Epidemiol*. 2013;37(6):603–13.
 54. Morrison JL, Gauderman WJ. GxEScanR: Run GWAS/GWEIS Scans Using Binary Dosage Files. R package version 2.0.2. 2020. <https://CRAN.R-project.org/package=GxEScanR>.
 55. Kraft P, Yen YC, Stram DO, Morrison J, Gauderman WJ. Exploiting gene-environment interaction to detect genetic associations. *Hum Hered*. 2007;63(2):111–9.
 56. Murcray CE, Lewinger JP, Gauderman WJ. Gene-environment interaction in genome-wide association studies. *Am J Epidemiol*. 2009;169(2):219–26.
 57. Kooperberg C, Leblanc M. Increasing the power of identifying gene x gene interactions in genome-wide association studies. *Genet Epidemiol*. 2008;32(3):255–63.
 58. Ionita-Laza I, McQueen MB, Laird NM, Lange C. Genomewide weighted hypothesis testing in family-based association studies, with an application to a 100K scan. *Am J Hum Genet*. 2007;81(3):607–14.
 59. Lewinger JP, Kawaguchi ES, Gauderman WJ. A note on p-value multiple-testing adjustment for two-step genome-wide gene-environment interactions scans. medRxiv. 2023.
 60. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*. 2008;17(R2):R122–128.
 61. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999;55(4):997–1004.
 62. Sun M, Bjorge T, Teleka S, Engeland A, Wennberg P, Haggstrom C, et al. Interaction of leisure-time physical activity with body mass index on the risk of obesity-related cancers: a pooled study. *Int J Cancer*. 2022;151(6):859–68.
 63. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Glied TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*. 2010;26(18):2336–7.
 64. Diez-Obrero V, Dampier CH, Moratalla-Navarro F, Devall M, Plummer SJ, Diez-Villanueva A, et al. Genetic effects on transcriptome profiles in colon epithelium provide functional insights for genetic risk loci. *Cell Mol Gastroenterol Hepatol*. 2021;12(1):181–97.
 65. Jordahl KM, Shcherbina A, Kim AE, Su YR, Lin Y, Wang J, Qu C, Albanes D, Arndt V, Baurley JW et al. Beyond GWAS of Colorectal Cancer: Evidence of Interaction with Alcohol Consumption and Putative Causal Variant for the 10q24.2 Region. *Cancer epidemiology, biomarkers & prevention* : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology 2022;31(5):1077–1089.
 66. Tian Y, Kim AE, Bien SA, Lin Y, Qu C, Harrison T, Carreras-Torres R, Diez-Obrero V, Dimou N, Drew DA et al. Genome-Wide Interaction Analysis of Genetic Variants with Menopausal Hormone Therapy for Colorectal Cancer Risk. *J Natl Cancer Inst*. 2022;114(8):1135–48. <https://pubmed.ncbi.nlm.nih.gov/35512400/>.
 67. Cohen AJ, Saiakhova A, Corradin O, Luppino JM, Lovrenert K, Bartels CF, et al. Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nat Commun*. 2017;8:14400.
 68. Lee J, Jolanki O, Kim D, Strattan JS, Kundaje A, Nordström K, Shcherbina A. ENCODE-DCC/atac-seq-pipeline: v1.9.1. 2020.
 69. Lee J, Strattan JS, Shcherbina A, Kagda M, Maurizio PL. ENCODE-DCC/chip-seq-pipeline2: v1.6.1. 2020.
 70. Su YR, Di CZ, Hsu L. Genetics, Epidemiology of Colorectal Cancer C: A unified powerful set-based test for sequencing data analysis of GxE interactions. *Biostatistics*. 2017;18(1):119–31.
 71. Tu L, Yan B, Peng Z. Common genetic variants (rs4779584 and rs10318) at 15q13.3 contributes to colorectal adenoma and colorectal cancer susceptibility: evidence based on 22 studies. *Mol Genet Genomics*. 2015;290(3):901–12.
 72. Yang H, Gao Y, Feng T, Jin TB, Kang LL, Chen C. Meta-analysis of the rs4779584 polymorphism and colorectal cancer risk. *PLoS One*. 2014;9(2):e89736.
 73. Kyu HH, Bachman VF, Alexander LT, Mumford JE, Afshin A, Estep K, et al. Physical activity and risk of breast cancer, colon cancer, diabetes, ischemic heart disease, and ischemic stroke events: systematic review and dose-response meta-analysis for the Global Burden of Disease Study 2013. *BMJ*. 2016;354:i3857.
 74. Morris JS, Bradbury KE, Cross AJ, Gunter MJ, Murphy N. Physical activity, sedentary behaviour and colorectal cancer risk in the UK Biobank. *Br J Cancer*. 2018;118(6):920–9.
 75. Peters U, Hutter CM, Hsu L, Schumacher FR, Conti DV, Carlson CS, et al. Meta-analysis of new genome-wide association studies of colorectal cancer risk. *Hum Genet*. 2012;131(2):217–34.
 76. Whiffin N, Hosking FJ, Farrington SM, Palles C, Dobbins SE, Zgaga L, et al. Identification of susceptibility loci for colorectal cancer in a genome-wide meta-analysis. *Hum Mol Genet*. 2014;23(17):4729–37.
 77. Tanskanen T, van den Berg L, Valimaki N, Aavikko M, Ness-Jensen E, Hveem K, et al. Genome-wide association study and meta-analysis in Northern European populations replicate multiple colorectal cancer risk loci. *Int J Cancer*. 2018;142(3):540–6.
 78. Rakshit S, Bhaskar LVKS. An Intergenic Variant rs4779584 Between SCG5 and GREM1 Contributes to the Increased Risk of Colorectal Cancer: A Meta-Analysis. In: Nagaraju GP, Peela S, editors. *Novel therapeutic approaches for gastrointestinal malignancies Diagnostics and Therapeutic Advances in GI Malignancies*. edn. Singapore: Springer; 2020. p. 159–69.
 79. Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet*. 2008;40(1):26–8.
 80. Derynck R, Akhurst RJ, Balmain A. TGF-beta signaling in tumor suppression and cancer progression. *Nat Genet*. 2001;29(2):117–29.
 81. Stabile H, Mitola S, Moroni E, Belleri M, Nicoli S, Coltrini D, et al. Bone morphogenic protein antagonist Dm/gremlin is a novel proangiogenic factor. *Blood*. 2007;109(5):1834–40.
 82. Hedjazifar S, Khatib Shahidi R, Hammarstedt A, Bonnet L, Church C, Boucher J, et al. The novel adipokine gremlin 1 antagonizes insulin action and is increased in type 2 diabetes and NAFLD/NASH. *Diabetes*. 2020;69(3):331–41.
 83. Kobayashi H, Gieniec KA, Wright JA, Wang T, Asai N, Mizutani Y, Lida T, Ando R, Suzuki N, Lannagan TRM et al. The Balance of Stromal BMP Signaling Mediated by GREM1 and ISLR Drives Colorectal Carcinogenesis. *Gastroenterology*. 2021;160(4):1224–1239 e1230.
 84. Ren J, Smid M, Iaria J, Salvatori DCF, van Dam H, Zhu HJ, et al. Cancer-associated fibroblast-derived gremlin 1 promotes breast cancer progression. *Breast Cancer Res*. 2019;21(1):109.
 85. Corsini M, Moroni E, Ravelli C, Andres G, Grillo E, Ali IH, et al. Cyclic adenosine monophosphate-response element-binding protein mediates

- the proangiogenic or proinflammatory activity of gremlin. *Arterioscler Thromb Vasc Biol.* 2014;34(1):136–45.
86. Seidah NG, Chretien M. Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain Res.* 1999;848(1–2):45–62.
 87. Mbikay M, Seidah NG, Chretien M. Neuroendocrine secretory protein 7B2: structure, expression and functions. *Biochem J.* 2001;357(Pt 2):329–42.
 88. Ziai J, Matloff E, Choi J, Kombo N, Materin M, Bale AE. Defining the polyposis/colorectal cancer phenotype associated with the Ashkenazi *GREM1* duplication: counselling and management recommendations. *Genet Res.* 2016;98:e5.
 89. Yusuf I, Pardamean B, Baurley JW, Budiarto A, Miskad UA, Lusikooy RE, et al. Genetic risk factors for colorectal cancer in multiethnic Indonesians. *Sci Rep.* 2021;11(1):9988.
 90. Jo Y, Yeo MK, Dao T, Kwon J, Yi HS, Ryu D. Machine learning-featured Secretogranin V is a circulating diagnostic biomarker for pancreatic adenocarcinomas associated with adipopenia. *Front Oncol.* 2022;12:942774.
 91. Farber CR, Chitwood J, Lee SN, Verdugo RA, Islas-Trejo A, Rincon G, et al. Overexpression of Scg5 increases enzymatic activity of PCSK2 and is inversely correlated with body weight in congenic mice. *BMC Genet.* 2008;9:34.
 92. Klimentidis YC, Raichlen DA, Bea J, Garcia DO, Wineinger NE, Mandarino LJ, et al. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including *CADM2* and *APOE*. *Int J Obes (Lond).* 2018;42(6):1161–76.
 93. Doherty A, Smith-Byrne K, Ferreira T, Holmes MV, Holmes C, Pulit SL, et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nat Commun.* 2018;9(1):5257.
 94. Fabre O, Giordani L, Parisi A, Pattamaprapanont P, Ahwazi D, Brun C, Chakroun I, Taleb A, Blais A, Andersen E et al. *GREM1* is epigenetically reprogrammed in muscle cells after exercise training and controls myogenesis and metabolism. *bioRxiv.* 2020:2020.2002.2020.956300.
 95. Wang T, Zhang Y, Taaffe DR, Kim JS, Luo H, Yang L, et al. Protective effects of physical activity in colon cancer and underlying mechanisms: a review of epidemiological and biological evidence. *Crit Rev Oncol Hematol.* 2022;170:103578.
 96. Dziewiecka H, Buttar HS, Kasperska A, Ostapiuk-Karolczuk J, Domagal-ska M, Cichon J, et al. Physical activity induced alterations of gut microbiota in humans: a systematic review. *BMC Sports Sci Med Rehabil.* 2022;14(1):122.
 97. Saeidi A, Seifi-Ski-Shahr F, Soltani M, Daraei A, Shirvani H, Laher I, Hackney AC, Johnson KE, Basati G, Zouhal H. Resistance training, gremlin 1 and macrophage migration inhibitory factor in obese men: a randomised trial. *Arch Physiol Biochem.* 2023;129(3):640–8. <https://pubmed.ncbi.nlm.nih.gov/33370549/>.
 98. Ataenosrat A, Saeidi A, Abednatanzi H, Rahmani H, Dalooi AA, Pashaei Z, et al. Intensity dependent effects of interval resistance training on myokines and cardiovascular risk factors in males with obesity. *Front Endocrinol (Lausanne).* 2022;13:895512.
 99. Pourteymour S, Eckardt K, Hohen T, Langleite T, Lee S, Jensen J, et al. Global mRNA sequencing of human skeletal muscle: search for novel exercise-regulated myokines. *Mol Metab.* 2017;6(4):352–65.
 100. Saran U, Guarino M, Rodriguez S, Simillion C, Montani M, Foti M, et al. Anti-tumoral effects of exercise on hepatocellular carcinoma growth. *Hepatol Commun.* 2018;2(5):607–20.
 101. Endo Y, Zhang Y, Olumi S, Karvar M, Argawal S, Neppi RL, et al. Exercise-induced gene expression changes in skeletal muscle of old mice. *Genomics.* 2021;113(5):2965–76.
 102. Entrez Gene: *KCNG1* potassium voltage-gated channel, subfamily G, member 1. Available at <https://www.ncbi.nlm.nih.gov/gene/3755>. In.
 103. Bray MS, Hagberg JM, Perusse L, Rankinen T, Roth SM, Wolfarth B, et al. The human gene map for performance and health-related fitness phenotypes: the 2006–2007 update. *Med Sci Sports Exerc.* 2009;41(1):35–73.
 104. Wong HL, Koh WP, Probst-Hensch NM, den Van Berg D, Yu MC, Ingles SA. Insulin-like growth factor-1 promoter polymorphisms and colorectal cancer: a functional genomics approach. *Gut.* 2008;57(8):1090–6.
 105. Ke J, Lou J, Chen X, Li J, Liu C, Gong Y, et al. Identification of a potential regulatory variant for colorectal cancer risk mapping to chromosome 5q31.1: a post-GWAS study. *PLoS One.* 2015;10(9):e0138478.
 106. Liu DX, Lobie PE. Transcriptional activation of p53 by Pitx1. *Cell Death Differ.* 2007;14(11):1893–907.
 107. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. *Int J Behav Nutr Phys Act.* 2008;5:56.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.